


Curso: ¿Cómo se maneja SPSS como un profesional? Introducción al trabajo con sintaxis

José Manuel Torrado
Profesor Ayudante Doctor
Departamento de Sociología, Universidad de Granada
josetr@ugr.es

LECTURA 2: EL TRABAJO CON DATOS SECUNDARIOS

Fuentes para el trabajo con datos secundarios

En la lectura 1 veíamos los pormenores del proceso de investigación social cuantitativo. Como ha podido observarse, el proceso que va desde la formulación de del objetivo hasta la obtención final de los datos implica un ingente trabajo de diseño debido a la necesidad de operacionalizar variables, diseñar el cuestionario, determinar la muestra, administrar el cuestionario y finalmente codificarlos. Llegados a este punto el alumno se preguntará si efectivamente los investigadores seguimos todos y cada uno de estos pasos a rajatabla cada vez que nos planteamos un objetivo de investigación, y si es así ¡pues vaya peñazo! La realidad es que sí, hacemos todo eso siempre queremos obtener datos primarios, es decir, datos obtenidos por el propio investigador específicamente para sus objetivos. Sin embargo, no siempre emprendemos ese largo y arduo camino, porque, por suerte, en nuestro país existen instituciones oficiales que producen datos secundarios que, aunque usualmente se diseñan con fines distintos a la investigación social, pueden ser útiles para responder a ciertos objetivos de investigación – la mayoría por suerte –. Los datos secundarios que producen estas instituciones nos ayudan a saltarnos todas las etapas del diseño metodológico de la encuesta y poder centrarnos casi exclusivamente en el proceso de análisis de datos. En nuestro país existen tres instituciones principales que producen datos públicos detallados a escala nacional: el CIS, el Instituto Nacional de Estadística (INE) y la agencia de estadística Europea (EUROSTAT), en la tabla 1 resumo un poco qué son estas fuentes y pongo algunos ejemplos de los datos que ofrecen.

Tabla 1. Algunas características de las tres principales instituciones productoras de datos públicos

Institución	Escala de los datos	Disponibilidad de datos simplificados (tablas)	Disponibilidad de los datos detallados (microdatos)	Ejemplos de fuentes más usadas en investigación social
INE	Nacional pero muchos con representatividad a niveles inferiores	Directa, sencilla, con opciones gráficas	Directa, pero proceso de descarga complejo	Padrón, Censo de Población, Encuesta de Condiciones de Vida
CIS	Nacional, excepto algunas encuestas autonómicas	Directa, sencilla	Directa, proceso de descarga fácil (directa en SPSS)	Barómetros mensuales
EUROSTAT	Internacional (Europeo)	Directa, sencilla, con opciones gráficas	A petición, previa solicitud	EU-SILC




Además, los datos secundarios ofrecidos por estas instituciones tienen el respaldo de las mismas, ofreciéndonos diseños metodológicos y muestrales sólidos, con muestras representativas a escala nacional, y con encuestas sobre diversidad de temas, que nos permiten estudiar un amplio abanico de procesos sociales – desde la salud (que nos ofrece por ejemplo la Encuesta Nacional de Salud del INE), hasta el comportamiento político (que nos ofrecen los tradicionales Barómetros del CIS) –. Por ello, los investigadores especializados en un tema, cuando nos planteamos un objetivo de investigación, antes de plantearnos iniciar todo ese arduo proceso de extracción de datos, lo primero que hacemos es buscar en esas fuentes de datos y, escarbando en ellas, buscar las encuestas publicadas que mejor se adaptan a los objetivos de investigación que planteamos.

Precauciones a tener en cuenta cuando usamos fuentes secundarias

Sin embargo, el trabajo con datos secundarios no está exento de problemas, es necesario tener en cuenta las limitaciones que presentan estos datos, las cuales a veces pueden ser escasas, pero en otros casos pueden ser muy grandes. De todas ellas, quisiera destacar dos que me parecen fundamentales.

En primer lugar, debemos siempre **considerar a qué escala nos permiten acercarnos los datos secundarios**. Para ello, tenemos que ver su ficha técnica en la que, entre otras cosas, se detalla el diseño muestral. La mayoría de las encuestas de las que dispone el CIS por ejemplo, son representativas a escala nacional, pero solo a escala nacional, por lo que no nos permitirían analizar tendencias que ocurren a nivel autonómico y, mucho menos, a nivel municipal. Por ello, antes de trabajar con datos secundarios, es conveniente comprobar de qué escala son representativos, para poder focalizarnos en ella y no extraer conclusiones apresuradas. Esta información está, como he dicho, en la ficha técnica.

En segundo lugar, tenemos que **confirmar que la encuesta que elegimos nos permite medir aquello que realmente nosotros queremos medir, y no otra cosa**. Para ello, tenemos ojear muy bien los cuestionarios que ha utilizado la institución productora de los datos para obtener la información, porque si no los hacemos, corremos el riesgo de malinterpretar las variables. Os pongo un ejemplo real. En mi grupo de investigación nos dedicamos a estudiar la movilidad residencial (el hecho de cambiar de residencia, las mudanzas vaya), y para ello, descubrimos que hay dos fuentes principales: la Estadística de Variaciones Residenciales y el Censo. Sin embargo, analizando con detalle ambas fuentes nos dimos cuenta de que contabilizaban como movilidad cosas distintas. Mientras la EVR contabiliza como mudanza de una persona a aquellos cambios que han supuesto un empadronamiento (es decir solo registra como movimiento de una persona el hecho de que se dé de baja en un municipio y se registre en otro; por ejemplo que se dé de baja en Huelva y se dé de alta en Granada), el Censo registra como tal el hecho de vivir en un sitio diferente al que se vivía hace 10 años. Ambas fuente miden cosas distintas, mientras la EVR registra un cambio administrativo, el Censo considera que te has mudado si tu residencia actual es diferente a la que tenías hace 10 años, y estas



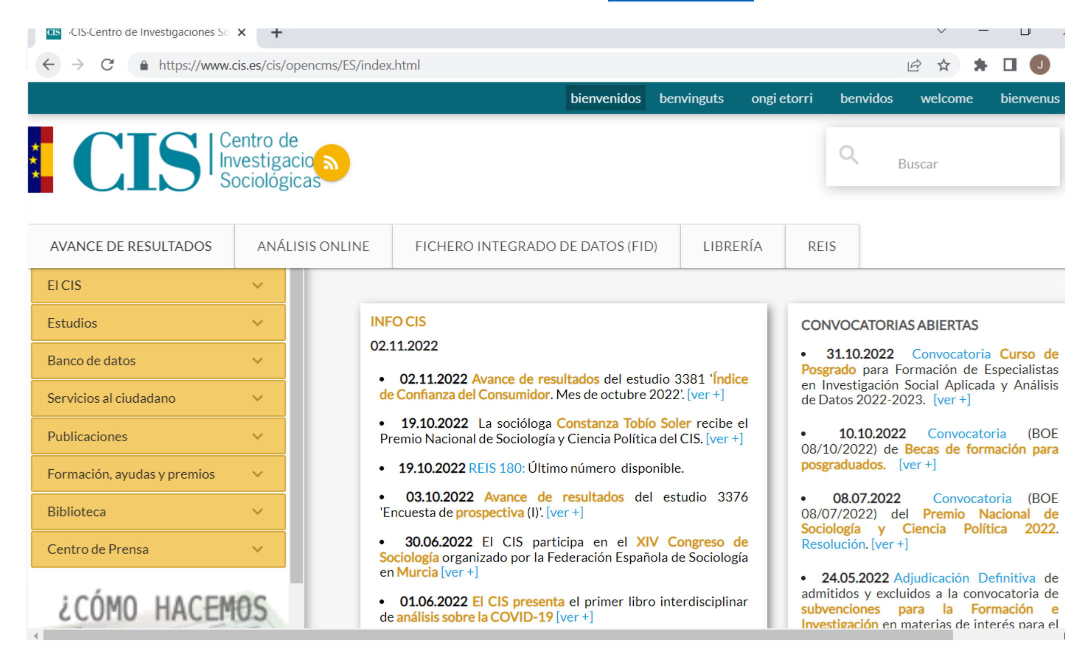
diferencias hay que considerarlas cuando analizamos el fenómeno empleando una fuente u otra.

Sobre el proceso de descarga de los datos secundarios

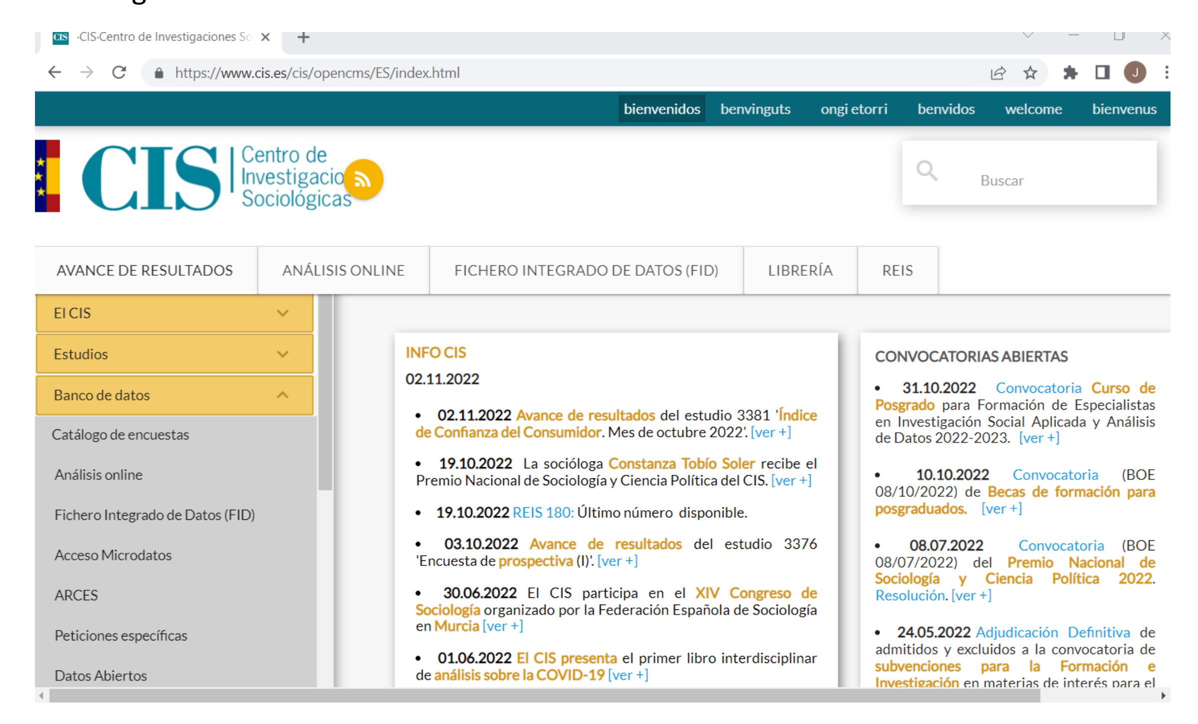
Más allá de las precauciones que señalo, también hay que considerar que el proceso de descarga de los datos, y especialmente su montaje, no siempre es sencillo. A veces las instituciones nos ofrecen los datos en dos archivos separados. Por un lado nos dan la "matriz de datos" en un archivo de texto (.txt) y por otro lado la "estructura de datos" en un archivo de Excel (.csv). Montar estos datos implica un proceso complejo de importación de la matriz de datos, y escritura de la "estructura de datos" en el lenguaje de programación de SPSS. Sin embargo, aunque este procedimiento sigue siendo el habitual en las bases de datos que ofrece el INE, en los últimos años el CIS ha mejorado y simplificado el proceso de descarga de datos, dando a día de hoy archivos directamente legibles con SPSS sin ninguna operación adicional por parte del usuario. Por ello, dado el carácter introductorio de este curso, en lo que queda de este documento os enseñaré cómo se descargan datos del CIS.

Descarga datos detallados (microdatos) del Centro de Investigaciones Sociológicas

Para descargar datos de encuestas del CIS, lo primero que tenemos que hacer es acceder a la web de dicha institución, en la dirección www.cis.es



Una vez ahí, clicamos en "Banco de datos", en la ventana de opciones que aparece a la izquierda de la pantalla. Y dentro de las opciones que se despliegan, la damos a "Catálogo de encuestas"



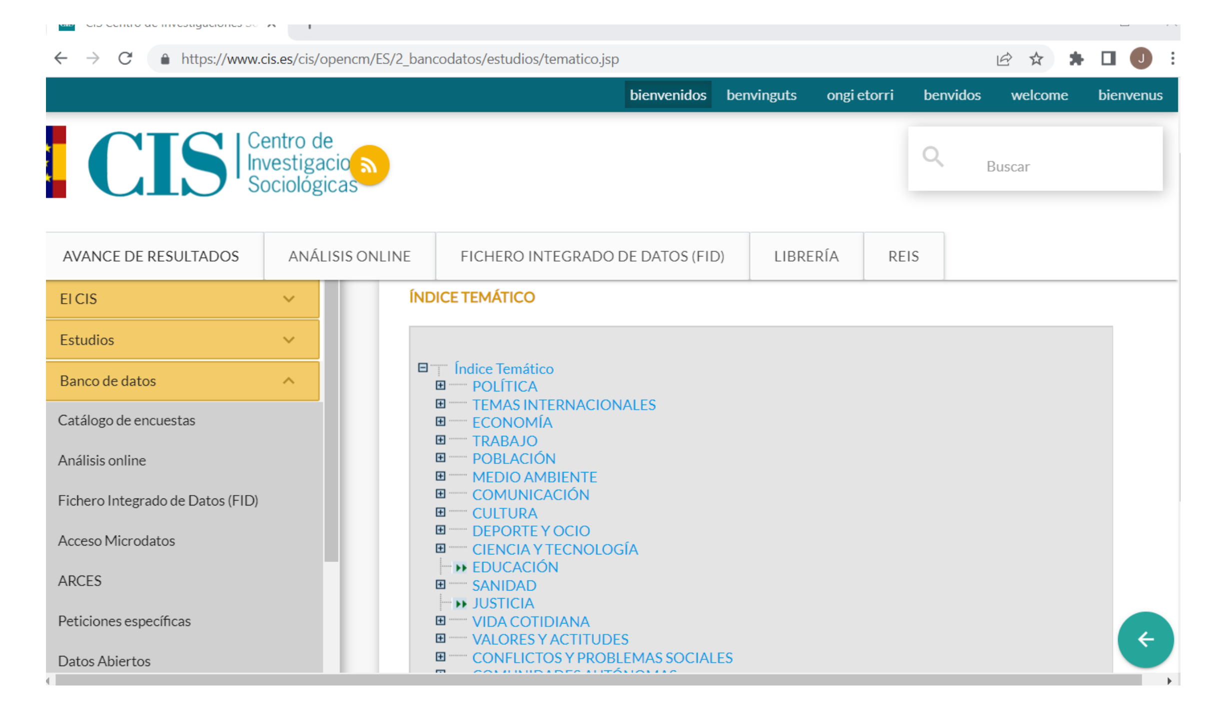
Accederéis a la ventana que veis abajo. Una vez ahí podéis optar por buscar encuestas siguiendo tres métodos "Índice temático", "Índice anual" y "Búsqueda asistida". Yo recomiendo el "Índice temático", ya que os permite elegir por temas, algo muy recomendable cuando tenéis claro un objetivo de investigación.







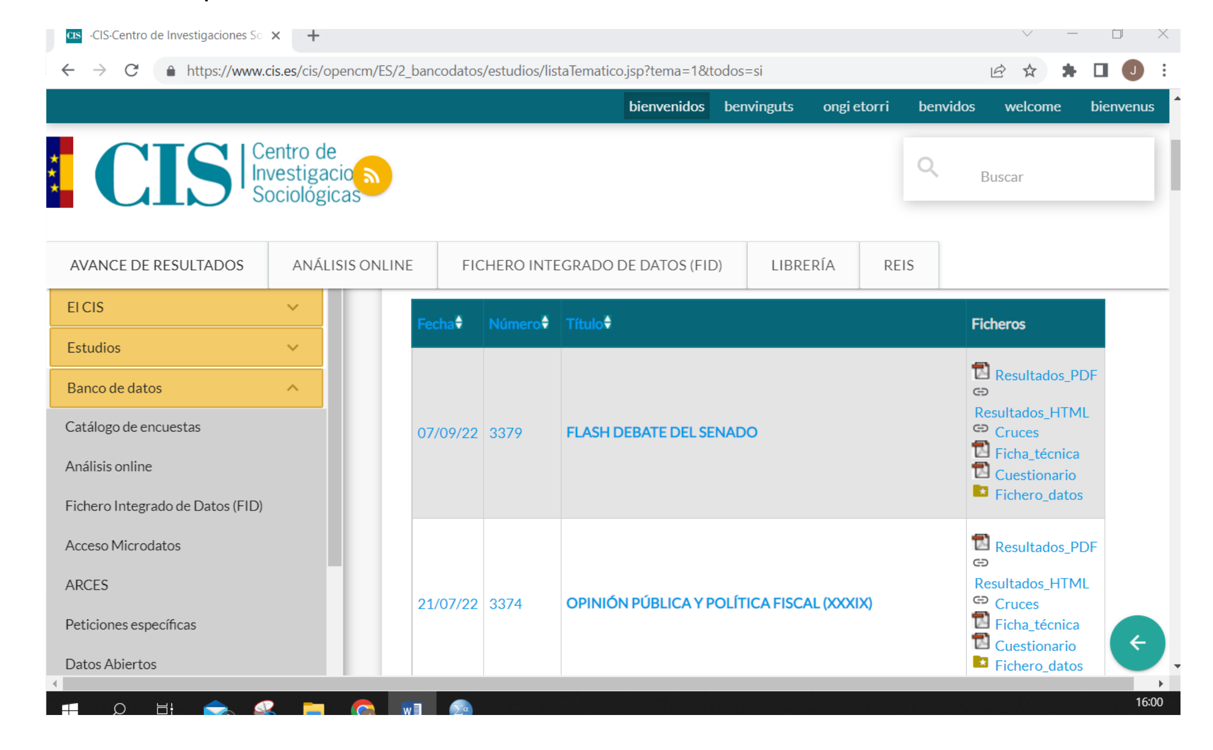
Una vez clicáis en "Índice temático" llegaréis a la siguiente ventana, en ella veis la variedad de temas de los que el CIS tiene encuestas descargables. Además, si clicáis en cada tema (símbolo "+") podéis acceder a distintos subtemas.



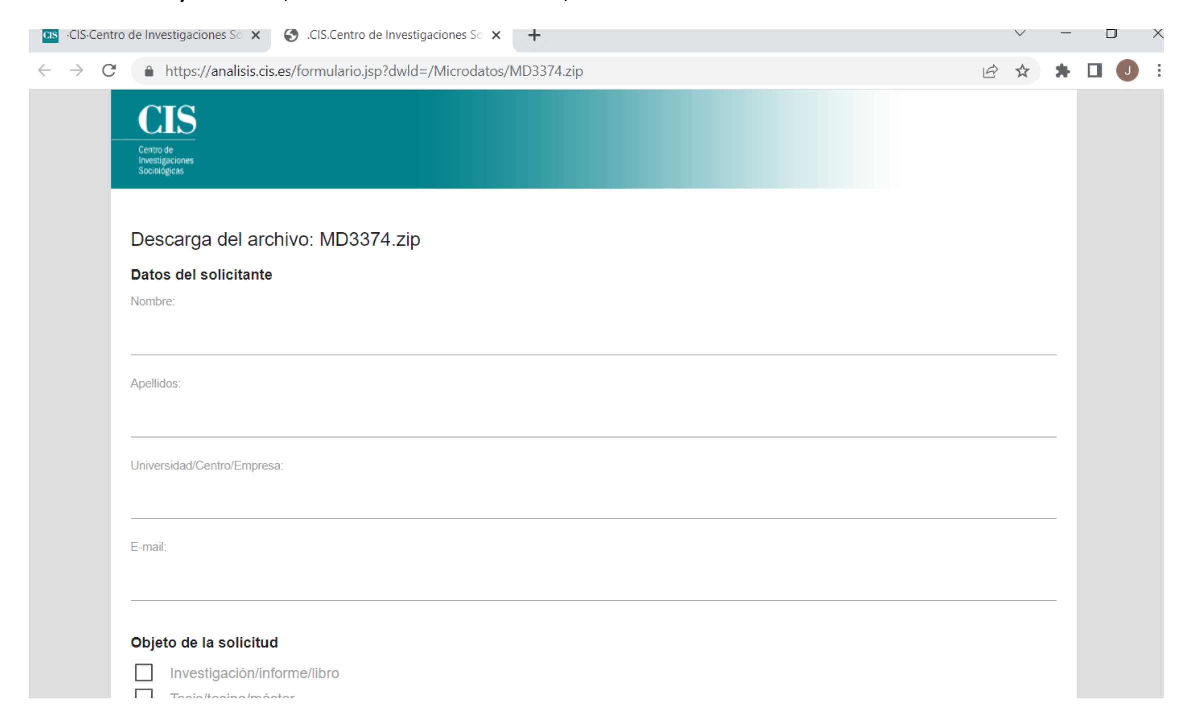


En este ejemplo he clicado directamente en la primera categoría "POLÍTICA", accediendo a la siguiente ventana. En ella podéis ver todas las encuestas que el CIS ha publicado sobre el tema ordenadas de más recientes a más antiguas. Si os fijáis, para cada encuesta os permite descargar 5 tipos de ficheros:

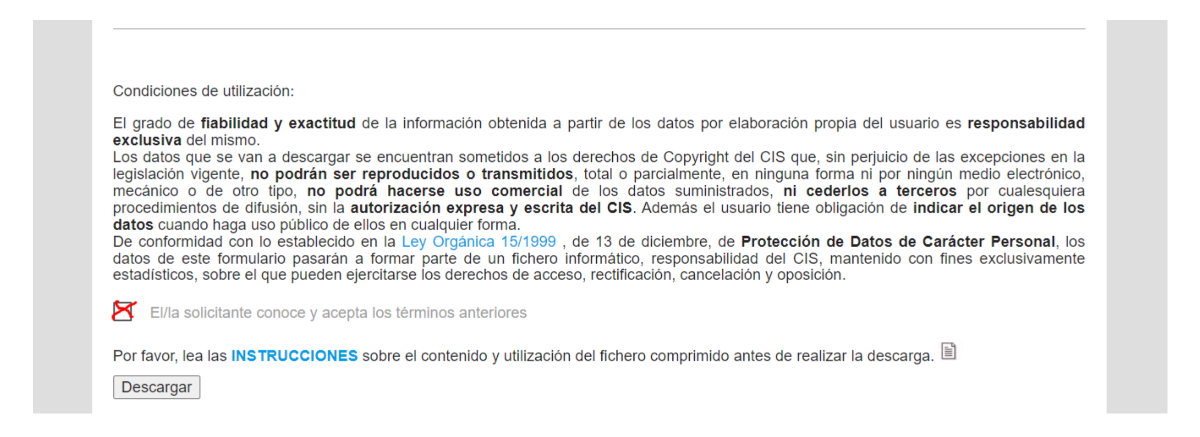
- Resultados_PDF, que es un análisis de frecuencias de todas y cada una de las preguntas del cuestionario.
- Resultado_HTML, que es lo mismo pero en versión web.
- Cruces_HTML, que son un conjunto de cruces de todas las variables del cuestionario según algunas variables clave (sexo, edad, ocupación, etc.) en versión web.
- Cuestionario, que es el cuestionario que el CIS ha empleado para elaborar la encuesta.
- Fichero_datos, que es un fichero de datos que, además de toda la información anterior, también contiene el archivo ejecutable en SPSS con toda la encuesta lista para su análisis.



Para descargar los datos de la encuesta clicad en “Fichero_datos” y se os abrirá la siguiente pantalla. Se trata de un formulario que emplea el CIS, es de uso meramente estadístico y formal, rellenadlo sin miedo, no os van a bombardear a SPAM.



Una vez relleno, recordad dar clic en “El/la solicitante conoce y acepta los términos anteriores” y le dais a descargar.



Condiciones de utilización:

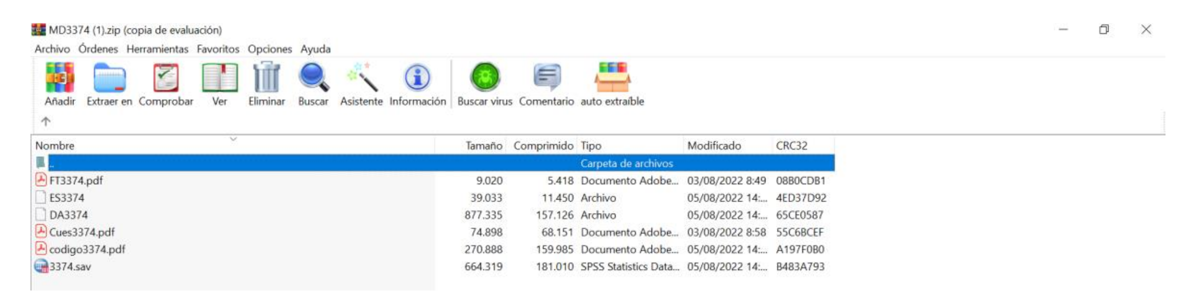
El grado de **fiabilidad y exactitud** de la información obtenida a partir de los datos por elaboración propia del usuario es **responsabilidad exclusiva** del mismo.
Los datos que se van a descargar se encuentran sometidos a los derechos de Copyright del CIS que, sin perjuicio de las excepciones en la legislación vigente, **no podrán ser reproducidos o transmitidos**, total o parcialmente, en ninguna forma ni por ningún medio electrónico, mecánico o de otro tipo, **no podrá hacerse uso comercial** de los datos suministrados, **ni cederlos a terceros** por cualesquiera procedimientos de difusión, sin la **autorización expresa y escrita del CIS**. Además el usuario tiene obligación de **indicar el origen de los datos** cuando haga uso público de ellos en cualquier forma.
De conformidad con lo establecido en la Ley Orgánica 15/1999 , de 13 de diciembre, de **Protección de Datos de Carácter Personal**, los datos de este formulario pasarán a formar parte de un fichero informático, responsable del CIS, mantenido con fines exclusivamente estadísticos, sobre el que pueden ejercitarse los derechos de acceso, rectificación, cancelación y oposición.

El/la solicitante conoce y acepta los términos anteriores

Por favor, lea las **INSTRUCCIONES** sobre el contenido y utilización del fichero comprimido antes de realizar la descarga.

Descargar

Se os descargará un carpeta en ZIP, si no podéis abrirla es porque no tenéis instalado ningún programa de descompresión de archivos ZIP, si es así, podéis instalar 7ZIP que es gratuito.

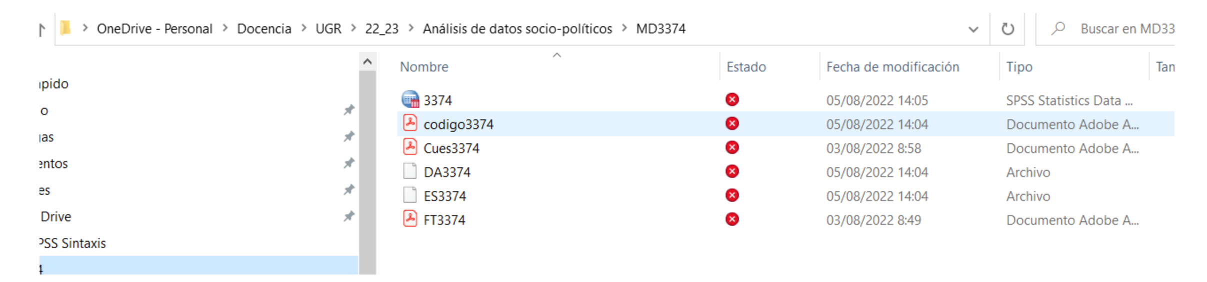


Nombre	Tamaño	Comprimido	Tipo	Modificado	CRC32
..			Carpeta de archivos		
FT3374.pdf	9.020	5.418	Documento Adobe...	03/08/2022 8:49	0880CD81
ES3374	39.033	11.450	Archivo	05/08/2022 14:...	4ED37D92
DA3374	877.335	157.126	Archivo	05/08/2022 14:...	65CE0587
Cues3374.pdf	74.898	68.151	Documento Adobe...	03/08/2022 8:58	55C68CEF
codigo3374.pdf	270.888	159.985	Documento Adobe...	05/08/2022 14:...	A197F0B0
3374.sav	664.319	181.010	SPSS Statistics Data...	05/08/2022 14:...	B483A793





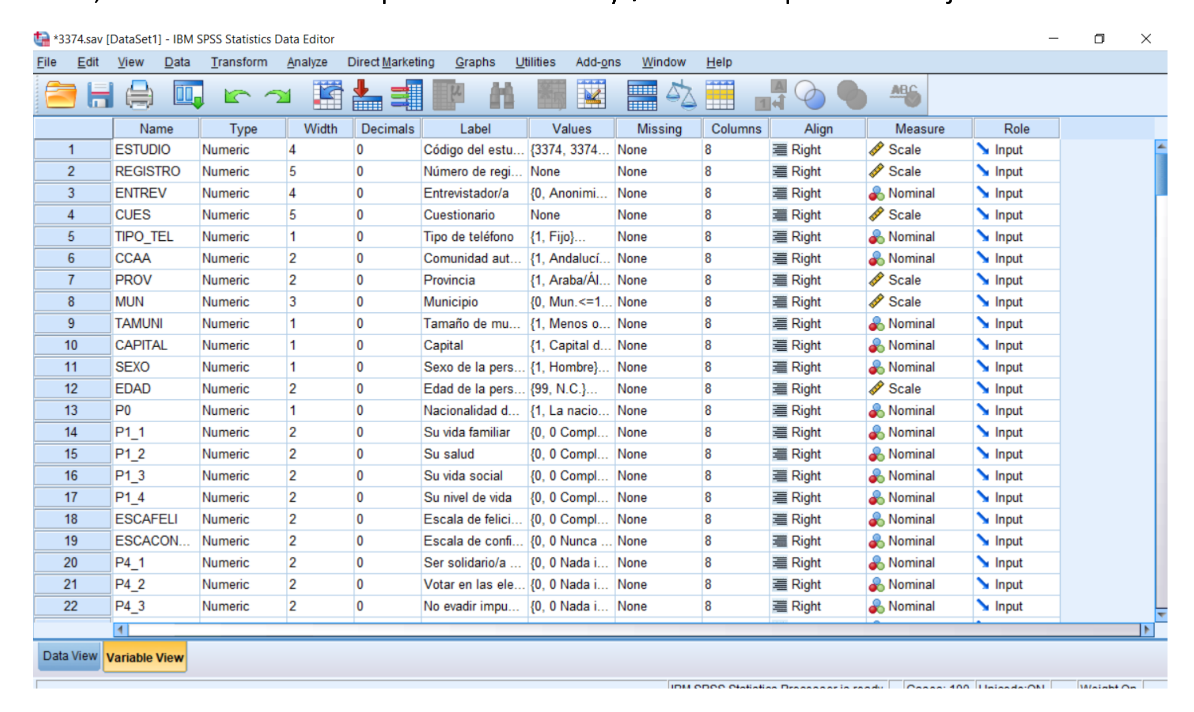
Una vez en el programa de descompresión, pedís que los archivos se descompriman en una carpeta que podéis llamar como queráis (sugiero el nombre de la encuesta de la que se trate). En la siguiente captura os muestro como quedaría.



OneDrive - Personal > Docencia > UGR > 22_23 > Análisis de datos socio-políticos > MD3374

Nombre	Estado	Fecha de modificación	Tipo
3374		05/08/2022 14:05	SPSS Statistics Data ...
codigo3374		05/08/2022 14:04	Documento Adobe A...
Cues3374		03/08/2022 8:58	Documento Adobe A...
DA3374		05/08/2022 14:04	Archivo
ES3374		05/08/2022 14:04	Archivo
FT3374		03/08/2022 8:49	Documento Adobe A...

El archivo numérico (en este ejemplo 3374) es el que directamente es ejecutable en SPSS, si lo tenéis instalado podréis darle clic y ¡tachán! Ya podéis trabajar con los datos



*3374.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ESTUDIO	Numeric	4	0	Código del estu...	{3374, 3374...	None	8	Right	Scale	Input
2	REGISTRO	Numeric	5	0	Número de regi...	None	None	8	Right	Scale	Input
3	ENTREV	Numeric	4	0	Entrevistador/a	{0, Anonimi...	None	8	Right	Nominal	Input
4	CUES	Numeric	5	0	Cuestionario	None	None	8	Right	Scale	Input
5	TIPO_TEL	Numeric	1	0	Tipo de teléfono	{1, Fijo}...	None	8	Right	Nominal	Input
6	CCAA	Numeric	2	0	Comunidad aut...	{1, Andalucí...	None	8	Right	Nominal	Input
7	PROV	Numeric	2	0	Provincia	{1, Araba/Ál...	None	8	Right	Scale	Input
8	MUN	Numeric	3	0	Municipio	{0, Mun.<=1...	None	8	Right	Scale	Input
9	TAMUNI	Numeric	1	0	Tamaño de mu...	{1, Menos o...	None	8	Right	Nominal	Input
10	CAPITAL	Numeric	1	0	Capital	{1, Capital d...	None	8	Right	Nominal	Input
11	SEXO	Numeric	1	0	Sexo de la pers...	{1, Hombre}...	None	8	Right	Nominal	Input
12	EDAD	Numeric	2	0	Edad de la pers...	{99, N.C.}...	None	8	Right	Scale	Input
13	P0	Numeric	1	0	Nacionalidad d...	{1, La nacio...	None	8	Right	Nominal	Input
14	P1_1	Numeric	2	0	Su vida familiar	{0, 0 Compl...	None	8	Right	Nominal	Input
15	P1_2	Numeric	2	0	Su salud	{0, 0 Compl...	None	8	Right	Nominal	Input
16	P1_3	Numeric	2	0	Su vida social	{0, 0 Compl...	None	8	Right	Nominal	Input
17	P1_4	Numeric	2	0	Su nivel de vida	{0, 0 Compl...	None	8	Right	Nominal	Input
18	ESCAFELI	Numeric	2	0	Escala de felici...	{0, 0 Compl...	None	8	Right	Nominal	Input
19	ESCACON...	Numeric	2	0	Escala de confi...	{0, 0 Nunca ...	None	8	Right	Nominal	Input
20	P4_1	Numeric	2	0	Ser solidario/a ...	{0, 0 Nada i...	None	8	Right	Nominal	Input
21	P4_2	Numeric	2	0	Votar en las ele...	{0, 0 Nada i...	None	8	Right	Nominal	Input
22	P4_3	Numeric	2	0	No evadir impu...	{0, 0 Nada i...	None	8	Right	Nominal	Input

Data View | Variable View

No obstante, y aunque las variables y sus categorías vienen etiquetadas, siempre recomiendo que miréis el cuestionario en caso de duda.